



EUROPEAN COMMISSION
EUROSTAT

Directorate F: Social statistics

DSS/2014/Sep/7.6

Item 7.6

**Vision Infrastructure Project (VIP) on streamlining and integration of
the European social surveys**

MEETING OF THE EUROPEAN DIRECTORS OF SOCIAL STATISTICS

LUXEMBOURG, 29-30 SEPTEMBER 2014

BECH BUILDING, ROOM QUETELET

1. PURPOSE OF THIS DOCUMENT

This document intends to provide an overview of the work done so far in the context of the Vision Infrastructure Project (VIP) “*Streamlining and integration of the European social surveys*” and to propose next steps.

The DSS is asked:

- to comment on the various strands of the work and on the proposed roadmap
- to express their priorities as regards actions to be tackled at EU level

2. BACKGROUND

In line with the 2009 Vision Communication¹, the Eurostat VIP project “*A limited-pillar approach to household/individual micro-level data collection for social statistics*”² was launched in 2011 with the aim of streamlining and integrating the European social surveys. The project has been supported since the beginning by the “Expert Group on the integration of the European social surveys” created by the DSS in 2011 with the mandate to address the issue of survey integration. It is composed by representatives of several NSIs.

Following the initial discussion with the Expert Group³, it was concluded that a more general approach would have to be adopted, not necessarily based on several pillars. Therefore, the project was renamed more generically as “*Streamlining and integration of the European social surveys*”.

Following the orientations endorsed by the DGINS in Wiesbaden in September 2011, the project is part of the overall programme for the modernisation of social statistics (see agenda item 7.1). At the same time, it contributes to the fulfilment of several of the strategic aims of the ESS Vision 2020⁴. In particular, it contributes to the objective “exploitation of existing databases and combination of multiple data sources to reduce response burden” by providing methodologies to combine data from the several European social surveys and to use the gain in efficiency to reduce the response burden.

The VIP “*Streamlining and integration of the European social surveys*” deals specifically with the second step of the process of modernisation of Social Surveys in the ESS. While the first step of this process is designed to introduce technical developments, adapt the current surveys to new policy needs, achieve a better coherence between surveys (through standardised variables and modules and integration across different domains), and also to develop framework regulation on social statistics, the second step of modernisation aims to create the

¹ *The production method of EU statistics: a vision for the next decade* COM(2009)404

² The project has also been referred in the past as “Pillar approach to social surveys” and “Integration of social statistics: Pillar approach”.

³ For more details, see document Eurostat/F/11/DSS/01/3.1EN of DSS meeting of 21-22 September 2011.

⁴ Final Report of Partnership Group’s Task Force Implementation of the ESS Vision 2020, presented at the DGINS Conference, 24 September 2014.

basis for a future fully integrated system of social surveys by reaping the methodological and technological possibilities that have emerged in the recent years. At the same time, the second step pursues the aim of accommodating resources constraints in the Member States through an innovative approach towards prioritisation and by setting the conditions to achieve further efficiency gains in the statistical production.

The project has been exploring technical and methodological solutions for an integrated and streamlined approach to micro-data collections for social statistics, including a possible roadmap towards their gradual implementation in the European social surveys. It includes all social (household/personal) surveys for which micro-data are transmitted to Eurostat: EU-LFS (Labour Force Survey), EU-SILC (EU Statistics on Income and Living Conditions), ICT-household Survey, AES (Adult Education Survey), EHIS (European Health Interview Survey), EHSIS (European Health and Social Integration Survey), HBS (Household Budget Survey) and HETUS (Harmonised European Time Use Survey).

The aim of the integration is to rationalise the current system of stand-alone social surveys by looking at possibilities for efficiency gains to be achieved so that the system of surveys becomes more flexible and more responsive, answering faster and being able to easily accommodate a wide range of new users' needs and policy demands.

Specifically, the aims are to:

- decrease the response burden (in terms of overall response burden and maximum burden that might befall a certain respondent);
- decrease the data collection cost;
- increase the potential for analysis;
- increase the flexibility of the European social surveys;

The approach followed in the project consisted so far in envisaging a fully modular survey architecture which is able to provide an answer to these several aims. This envisaged survey architecture was then used as a reference to identify methodological developments and actions leading to higher levels of integration in a coherent way having a target in mind. Based on the methodological developments and the actions identified, a preliminary roadmap for their adoption was designed. In the following steps of the project, these methodological developments, actions and corresponding roadmap will be discussed and further improved by the Expert Group.

The integration of the European social surveys can then be based on a progressive discussion, piloting and eventual adoption of the methodological developments and actions along the lines of the roadmap. The adoption of the developments will then progressively move the system of European social surveys in the direction of the envisaged architecture towards higher levels of integration.

3. METHODOLOGICAL DEVELOPMENTS

3.1. Initial phase

During the initial phase of this project various approaches (pillar approaches i.e. one to three modular surveys) to integrate the European social surveys were considered. The Expert Group on the integration of the European social surveys met three times to discuss alternative

approaches as well as the various approaches currently developed/implemented in several Member States. The modular approach was identified as the one eligible for in depth investigation and the modularization the necessary step for the streamlining and reorganisation of the system of surveys.

In its last meeting of February 2013 the Expert Group reviewed the concept of modular architecture, the advantages of reusing the data collected on those modules which are used in more than one survey and the targeted design of a system of interlinked micro-datasets. The current system of social surveys could be thought as a system of micro-datasets each of which corresponding to one of the current surveys. A modular architecture would introduce a higher degree of flexibility in terms of joint distributions and allow a higher degree of integration between different datasets. The system could be envisaged as consisting of a large number of shorter micro-datasets linked by core modules introduced in every survey instrument and by common modules standardised across statistical domains where appropriate.

The Expert Group supported the approach from a methodological point of view but recognised that considerable work needed to be done in order to arrive at a concrete proposal for the European social surveys, illustrating efficiency and costs gains and taking practical implementation issues into account.

3.2. The concept of modularity

The modular architecture of social surveys is based on the organisation of the content of the surveys in modules. In such architecture the modules are mutually exclusive groups of micro-data variables. The modules would in turn be distributed into survey instruments in such a way that each instrument consists of a fixed set of modules and each standard module may be present in one or more instruments. The social surveys would then consist, as they are today, of one or more instruments (with each instrument defining a particular micro-dataset to be sent to Eurostat) which provide information on several modules from the same statistical unit.

This type of architecture is based on the re-usability of the standard modules (building block of the system) in the development of different survey instruments. It impacts all the steps of the statistical production and the main potential efficiency gain (and challenge) comes from the impact it is expected to have on the data collection process. Standard modules would allow the use of standard questionnaire modules, standard data processing routines, standard interviewer training material, and so on, in different surveys, in different statistical domains.

Data compilation for each instrument may take place independently of the other instruments or with different degrees of coordination with them. The estimation of target parameters would pool the required input data from all instruments with which they have been compiled (a module may be present in more than one instrument). This is expected to produce benefits in the form of reduced response burden, increased precision and increased analytical potential of the data. In addition even estimates regarding variables that are present in only one instrument can benefit from such a pooling of information if they are strongly correlated to the variables which are common to more than one instrument, provided that the estimation method is appropriately chosen to take advantage of this pooling.

Moreover thanks to the deconstruction of the social surveys into modules the “stovepipe” nature of the current social survey would be replaced by an integrated and more flexible system of surveys which leaves the possibility to ‘move’ modules between instruments or to add new modules in the system. This allows the system to respond quickly and efficiently to emerging needs for statistical information.

3.3. Methods for survey integration

In the context of the VIP project “Streamlining and *integration of the European social statistics*” Eurostat has been trying to improve the understanding of how the future architecture of a system of interlinked micro-data collections could look like and how it could be achieved.

The first methodological studies run by Eurostat in 2012 consisted of a literature review on sampling design and estimation in the context of integrated systems of social surveys and have been already presented to the Expert Group.

Methodological contracts specifically focused on the development of methods and scenarios for an integrated system of European social surveys have run during 2013 and 2014. The objective was to study which methodological developments can be used to increase the overall efficiency and flexibility of a system of surveys and what implications it might have for the European system of social surveys.

The results of these studies contain methodological and practical recommendations for the organisation of integrated social surveys in the European Statistical System as well as practical issues to be addressed in the organisation of a pilot implementation of an integrated system of surveys.

The main results of the methodological developments of the study include:

- **Estimation methods** which may be applied in the context of the integration of surveys for pooling data. They include algorithms for combining or for aligning different estimates of the same quantity originating in different instruments;
- **Sample size determination methods for modules**, designed to guarantee that estimates of a quantity, possibly estimates combined from information in different instruments, achieve a certain total (across instruments) sample size and therefore a certain given precision;
- **Sample size determination methods for instruments**, designed to optimise the choice of the sample size of an instrument, so that all modules, but also all groups of modules, for which joint presence in an instrument is demanded (mandatory crossings), are administered to a sufficient number of sample units;
- **Instrument composition methods** to determine the optimal composition in terms of modules for each frequency, so as to minimise response burden while satisfying certain constraints concerning the concordance of frequencies, the joint presence of certain modules in some instrument, etc.

Another important output of the methodological studies is the specification of a general and provisional road map (list of actions) for the progressive implementation of these methods in the direction of a future integrated system of European social surveys (see section 3.5).

3.4. Testing of the methodological developments

The methodological developments have been the object of a simulation study applied to the LFS, EU-SILC and AES surveys. The statistical procedures were programmed⁵ and tested in various alternative scenarios for the composition of the instruments having in mind the general aims of the integration of the European social surveys.

⁵ The statistical procedures were programmed in R (<http://www.r-project.org/>).

In order to apply the algorithms, LFS, EU-SILC and AES variables were split into 50 mutually exclusive modules, each one including 10-15 variables. The scenarios were different with respect to the number of instruments and to the characteristics of the instruments. While some fixed characteristics (free parameters) were common for most scenarios - frequency, legal precision requirements, average burden in terms of variables included in the modules, dependencies between modules, thematic groups of modules, presence of certain modules in all or certain instruments – other parameters were free in order to generate the alternative scenarios (number of instruments for each frequency of interest, precision requirements for modules not covered by legal restrictions, set of mandatory crossings).

For each scenario resulting from altering one or more of the free parameters, the optimisation problem consisted of searching the optimal instrument composition and the sample size allocation across instruments which could minimise the total cost (simplified function of the number of respondents and burden of the modules).

Seven alternative scenarios with different options for the free parameters were examined, each one increasing the flexibility (as for example assuming frequency differentiation across modules, higher precision differentiations, etc.) and/or making information requirements more demanding (increasing the demand of information on joint distributions in terms of mandatory crossings). The benchmark scenario to which alternatives are compared was defined as close as possible to the current split in LFS, EU-SILC and AES.

The results of this first simulation showed that the statistical procedures were able to find solutions (i.e. instrument compositions) which would be less costly while complying with the constraints specified for each of the scenarios. In particular:

- cost/burden reduction could be achieved by increasing flexibility, especially when this was done in a way that one takes advantage of synergies between the various flexibility aspects: increasing the sample size differentiation may bring cost reductions, when the number of instruments is allowed to vary in parallel;
- certain simple steps towards more flexible solutions, as e.g. allowing some frequency differentiation across modules, allowed the procedures to find burden/cost reductions;
- increasing the demands on the statistics to be produced increases the burden but the procedures could still find ways to satisfy such needs for information demanded by users while still achieving important burden reductions as compared to the initial benchmark scenario.

The next step is to assess more thoroughly, and possibly quantify, how these methodological developments can deliver lower response rates, lower data collection costs, higher potential for analysis (i.e. possibility to cross data from different statistical domains) and more flexibility and responsiveness to new users' needs, in the particular context of the European social surveys.

3.5. Provisional roadmap

The actions which may be needed to gradually achieve an integrated modular architecture (covering both the European and the national levels) in the second step of the modernisation process may be organised into four streams:

- **Organisation:** actions that set up the organisational structures and procedures at European and national level (that will assess user needs, will implement the modular system, will

investigate its performance and will revise its specifications in time for each programming period).

- **Content:** actions required for the specification of harmonised variables and modules.
- **Methodology:** actions that deal with the further development and dissemination of survey and statistical methodology needed in order to implement modular surveys⁶ and analyse the data compiled by them.
- **IT infrastructure:** actions that cover the specification, design and development of software and standards that will support the data sharing, storing and processing needs of a European modular statistical production system. This stream sets goals for the development of a complete, integrated IT system, which is a lengthy process. The system however consists of three interoperable, modular sub-systems each one of which implements part of the methods and procedures needed under modularity. The gradual implementation of the sub-systems enables the gradual streamlining of the European Social Surveys in parallel with the three other streams (organisation, content and methodology).

In this draft formulation the actions foreseen in each stream follow a logical sequence. Each one provides input or creates the condition needed for implementation of subsequent actions on the same stream or in other streams. The system is gradually “maturing” towards a complete modular architecture according to a sequence of three maturity levels which may represent the requirement for changes to the current system social surveys:

- **1st level of maturity (short-middle term horizon):** The system allows the pooling of data from different surveys for the computation of estimates according to the estimation procedures (based on combining estimates or aligning the data from different surveys) which have been developed for point and precision estimates. The harmonisation of social variables is the main requirement.
- **2nd level of maturity (middle term horizon):** Additionally to the previous level of maturity, the system allows to modify the sample size of each survey if the precision requirement set for indicators estimated for its data can be achieved by combining data with those for other surveys (sample size can be shifted between surveys which share common variables and no actual modularization is required). The main requirement is the revision and harmonization of precision requirements so that they are tied to indicators (and not to indicators derived from a specific surveys)
- **3rd level of maturity (long term horizon):** the complete modularization of the system (composed of inter-linked micro-data sets). It represents the ultimate aim of the social survey integration project, the target in the long run. The envisaged system of modular surveys is operational at national and Eurostat level.

The path to a fully integrated system is difficult, requiring investment, change in the statistical production and uncertainty. The draft road map has to be considered with caution as it is necessarily a lively document which will need to be updated according to new events or relevant findings during the course of the modernisation. The first two levels of maturity do not require the complete modularisation as such but rather foster the harmonisation across variables and systems in order to accommodate the pooling of data. In this case the instruments can be considered similar to the current surveys but precision gains and/or sample size savings can still be achieved.

⁶ ‘Surveys’ in the broad sense, encompassing also collections of data from administrative sources.

An attempt to provide a more specific list (draft) of actions which can be envisaged for each stream is presented in Table 1.

4. FUTURE EUROPEAN SOCIAL SURVEYS ARCHITECTURE

The implementation of the methodological developments and actions until the 3rd level of maturity would result in an integrated survey architecture in which the system of modules would be designed on the basis of precise output requirements and users' needs. This includes:

- (1) The definition of the European data requirements (current users' priorities) in terms of modules and frequency at which they have to be implemented. The output have to be linked to modules or to a restricted number of modules (as these modules will constitute the building block for the future modular architecture) instead that be linked to complete surveys;
- (2) The definition of the European data requirements in terms of joint distributions (real users' needs) between modules. The crossings have to be linked to groups of two or more modules that contain variables whose joint distribution must be observed;
- (3) The definition of the European data requirements in terms of methodological characteristics for indicators, modules and joint distributions. Requirements as precision, frequency, longitudinal component, etc. can be analysed systematically taking into account the respective users' needs. The reduction of unnecessary requirements as regards the precision, frequency and/or longitudinal component for certain joint distributions would increase the flexibility of implementation for the Member States.

Eurostat thinks that such an approach would have positive consequences on the size of individual survey instruments, would reduce burden, facilitate the move towards data warehouse approaches, and would provide further flexibility to national compilers, among others.

5. NEXT STEPS

The activities envisaged in the roadmap need to be discussed and further refined by the DSS Expert Group on the integration of the European social surveys.

It is foreseen to submit the results of the methodological work carried in 2013-2014 on the technical possibilities available to deal with the challenges of an integrated system of surveys to the Expert Group in order to discuss the implication of those results for a future architecture of European social surveys. A meeting of the Expert Group might be organised in the first quarter of 2015.

Methodological work to deal with the challenges of an integrated system is expected to continue in 2015 with a focus on the implementation of the methods developed to show the efficiency gains which can be obtained by pooling of data. To this respect a study on the potential gains of adopting a modular surveys architecture for the European social surveys will be carried out and the result presented to the Expert Group.

Table 1 – Actions of the roadmap of modularization of the European social surveys

Organisation						
O1. Stakeholder consultation	O2. Mapping of the current national system of social surveys.	O3. Assessment of the modularisation options available under the current system	O4. IESS framework regulation	O5. Frame alignment in the current social surveys.	1 st level maturity	
	O6. Frame alignment in the current social surveys.	2 nd level maturity				
	O7. Revision of the sample sizes of the current social surveys so as to achieve precision requirements in an optimal manner.	O8. Design the business architecture at national level.	O9. Design the business architecture in Eurostat.	O10. Define coordination bodies and procedures for the specifications of each programming period.	O11. Definition administrative procedures for specifying the composition of instruments.	3 rd level maturity
Content						
C1. Harmonisation of definition of social variables. And indicators	1 st level maturity					
C2. Specification of the precision requirements of the harmonised indicators.	2 nd level maturity					
C3. Actual specification of the modules of social surveys.	C4. Specification of user needs at European (crossings, required precision, frequency).	3 rd level maturity				
Methodology						
M1. Accommodate longitudinal modules in the modular system of surveys.	M2. Investigate further point estimation and variance estimation methods (based on pooling data, for complex indicators)	M3. Specify the input information needed for the implementation of data pooling between surveys.	1 st level maturity			
M4. Specify the input information needed for the implementation of sample size re-allocation between surveys.	2 nd level maturity					
M5. Improve the current instrument composition algorithm.	M6. Specify the input information needed for the implementation of the instrument specification procedure.	M7. Investigate pros/cons of optimising instrument composition (national and European perspectives)	M8. Develop guidance for the assessment of contextual and mode effects.	3 rd level maturity		
Infrastructure (IT)						
I1. Integrate the needs of the modular system for the transmission of information between national authorities and between them and Eurostat into the ESS Information Models and Standards.	I2. Develop software system for data pooling between surveys and the methods developed for point and variance estimation	1 st level maturity				
I3. Develop an interoperable software module for sample size re-allocation between surveys and integrate it with I2	2 nd level maturity					
I4. Develop an interoperable software module for instrument composition specification and integrate it with I3.	3 rd level maturity					